Predicting Pathways Between Distant Configurations for Biomolecules

Konstantin Röder* and David J. Wales*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

E-mail: kr366@cam.ac.uk; dw34@cam.ac.uk

Abstract

Many of the most interesting rearrangements associated with function and dysfunction of biomolecules involve complex, highly non-linear pathways. Predicting these convoluted changes in structure is an important research challenge, since knowledge of key intermediate conformations at an atomic level of detail has the potential to inform the design of novel therapeutic strategies with enhanced specificity. The identification of kinetically relevant pathways can be strongly dependent on the construction of a physically relevant initial pathway between specified endpoints, avoiding artefacts such as chain crossings. In this contribution we describe an enhanced interpolation procedure to characterise initial pathways for complex rearrangements of a histone tail, α-helix to β-sheet conversion for amyloid-β_{17-42}, and egfr kinase activation. Complete connected initial pathways with relatively low overall barriers are obtained in each case using an enhanced quasi-continuous interpolation scheme. This approach will help to extend the complexity and time scales accessible to computer simulation.

*To whom correspondence should be addressed

1 Introduction

The potential energy landscape of a molecule encodes all the information necessary to understand the thermodynamic, kinetic, and structural properties.[1–5] Hence efficient methods to explore this landscape represent an active research field.[6]

Biomolecules such as proteins, RNA and DNA are the targets of many theoretical and experimental studies due to their central role in the chemistry of life and disease. However, the exploration of the potential energy landscape for some of the most interesting biomolecules is challenging, simply because of their size and the complex folding pathways involved.

In experiment and simulation a multitude of relaxation time scales may be observed for biomolecules if the time resolution is short enough. These time scales result from the large range of barrier heights on the landscape, where the highest barriers are usually associated with the slowest rate determining steps that determine function. The presence of high barriers can lead to broken ergodicity, and consequently enhanced sampling methods are generally required in simulations to provide experimentally relevant results.

The computational potential energy landscape framework uses geometry optimisation methods to locate local minima, and the transition states that connect them, to construct discrete paths,[7,8] which form a kinetic transition network.[9,10] Geometry optimisation can locate transition states efficiently almost independent of the associated barrier heights, which enables all the relevant time scales to be accessed by post-processing using the tools of statistical mechanics and unimolecular rate theory.

A key initial step in the construction of kinetic transition networks is the characterisation of a physically relevant pathway between products and reactants. The double-ended procedures that are usually employed for this task generally involve a series of images in configuration space that interpolate between two initial configurations. Our usual approach employs the doubly-nudged[11] elastic band[12,13] algorithm to suggest transition state candidates. These candidates are then accurately converged to transition states using hybrid eigenvector-following.[12,14,15] The simplest way to initialise the images in the band is a linear

interpolation. However, this procedure can introduce unphysical geometries, which subsequently lead to a variety of problems.

The three key, unphysical changes, which we have often observed for initial interpolations between distant configurations of large biomolecules, are chain-crossings of bonds (backbone and side-chains), chirality inversions, and cis-trans isomerism. These changes are usually associated with relatively high energy barriers, but the corresponding transition states can overlap in energy with physically relevant pathways, making it difficult to remove them systematically. Ideally, our interpolation would explicitly avoid such states, but post-sampling detection has often been required to deal with such problems.

A new scheme, quasi-continuous interpolation (QCI), was recently described,[16] which uses discrete images coupled with penalty terms in the band energy, to prevent chain crossing. The starting point for the QCI algorithm is the construction of a percolating network of constraints, which are used as an auxiliary potential for the atom-by-atom construction and refinement of the band images. Chain crossings can largely be avoided if a diverse database of valid configurations (usually local minima) is used to construct the auxiliary potential. However, it can be challenging to define all the covalent bonds and constrain them properly using a reference database. Incorrectly constraining non-bonded contacts that actually need to change prevents a physically realistic pathway from being found.

In the present contribution we use sequence information to construct the constraint potential unambiguously. We further demonstrate that the addition of constraints to restrain peptide bond dihedral angles allows for the explicit conservation of cis-trans isomers. Benchmark results are presented, optimising input parameters for pathways corresponding to the α-helix to β-sheet conversion for an amyloid-β_{17-42} monomer, a transformation pathway for a histone tail, and the location of low energy transition states for egfr kinase activation.

The QCI algorithm is related to the geometric targeting (GT) approach,[17,18] which aims to create plausible all-atom pathways. Both algorithms are based on geometric constraints and GT also employs stereochemical information. However, QCI is used to initialise the

location of transition states. The use of discrete images plus a quasi-continuous term in the auxiliary potential distinguishes QCI from continuous path methods.[19–23] The key difference from targeted molecular dynamics,[24,25] string methods[26–28] and other molecular dynamics based approaches, for example nebterpolation,[29] is the use of an inexpensive interpolation potential to seed the more expensive pathway calculations.

The example systems were selected to be biologically relevant, as well as illustrative of typical pathways associated with function in biomolecules. The Aβ_{17-42} monomer provides an ideal test system to optimise algorithmic parameters. This peptide is intrinsically disordered,[30] which means there is significant structural heterogeneity, with high energy barriers between alternative structures. While understanding the behaviour of the peptide and its aggregates is essential for research into Alzheimer's disease,[31–34] it has proved difficult in experiment and simulation to obtain insight into structure and dynamics.[35,36] These features produce a complicated, but important, transition pathway for benchmarking.

Histones are a protein component of chromatin, functioning as spools for DNA winding. Positively charged regions within the protein, known as histone tails, mediate interactions within the DNA-protein complexes.[37–40] The key aspect of the histone tail conformations considered for testing the QCI algorithm is the global rearrangement of the backbone, requiring the tightly packed side chains to rearrange as well. These complex motions lead to a large number of unphysical structures for naive interpolations between the endpoints.

Our last example is egfr kinase, which has been identified as a primary target for cancer treatment.[41] Understanding the activation process could provide new avenues to treat disease. This system has previously been used to test novel sampling strategies,[42] providing a useful reference for comparison with QCI.

Each test system constitutes a relatively large biomolecular machine, and the number of degrees of freedom increases the computational expense significantly. Furthermore, the size means that even for the permutationally best aligned structures the distances that must be bridged in the initial interpolation are large. The three examples constitute a representative

set of problems encountered in contemporary computational studies of biomolecules.

Our results demonstrate that these complex rearrangements can be characterised efficiently at an atomic level of detail. The QCI approach facilitates access to pathways associated with important function and malfunction with a wealth of potential applications.

2 Methods

2.1 Setup for connection attempts

For all the present simulations the AMBER[43] ff14SB[44] all-atom force field was used, properly symmetrised[45,46] as for the ff99SB[47] force field. An implicit Generalised Born solvation model (igb=2)[48,49] with the Debye-Hückel approximation for salt (0.1 M)[50] and infinite cutoffs was applied. For the egfr kinase, GPU acceleration was exploited.[51–53]

The two endpoints for the Aβ_{17-42} monomer were taken from two distinct funnels on the potential energy landscape computed within the discrete path sampling framework.[7,8] A full report on the exploration of this landscape is in preparation and will be presented elsewhere. The histone tail conformations were taken from replica-exchange molecular dynamics simulations (REMD).[54] The structures represent different minima from two-dimensional projections derived from a dihedral principal component analysis. The egfr end points are the active conformation from PDB entry 2ITW[55] and the inactive configuration from PDB entry 3W32.[56] The terminal regions were shortened to produce consistent lengths for this protein, yielding a sequence of 284 amino acids starting at residue number 701. The two endpoints are illustrated for all three systems in Fig. 1.

2.2 Summary of the original algorithm

The QCI algorithm[16] is designed to provide physically acceptable, low energy configurations that are used to initiate transition state calculations and construction of a kinetic transition network. The algorithm needs to be combined with a sampling scheme, for example

discrete path sampling,[7,8] as in the present work. The algorithm consists of three parts: the definition of an interpolation potential, construction of the interpolation images with energy optimisation, and finally minimisation of the RMS force on the band of images.

A key feature of the original QCI algorithm is the introduction of an auxiliary interpolation potential. For all bonded atom pairs a harmonic constraint is applied outside a given threshold r_{attr}, allowing the bonds to fluctuate around the observed equilibrium value. For all non-constrained pairs of atoms a Coulombic repulsion term is introduced if the atoms are within a cutoff distance r_{rep}. An additional repulsive term is introduced for any local minima in the distance between pairs of atoms in adjacent images to add a worst case energetic penalty. This last term is the quasi-continuous part of the potential.

The full auxiliary potential is then defined over all interpolation images, $\mathbf{X}_i$, with $i = 1$ and $i = N$ being the endpoints, as

$$V_{\text{aux}} = \sum_{\text{images}} V_{\text{constr}} + \sum_{\text{images}} V_{\text{rep}} + V_{\text{qc}}, \tag{1}$$

where the constraint potential is

$$V_{\text{constr}} = \sum_{a,b \in \text{constr}} V_c\left(a,b\right), \tag{2}$$

$$\text{with } V_c(a,b) = \begin{cases} 0, & \text{if } \mid r_{ab} - \bar{r}_{ab} \mid < r_{attr}, \\ k\dfrac{\left[(r_{ab} - \bar{r}_{ab})^2 - r_{attr}^2\right]^2}{2r_{attr}^2} & \text{otherwise;} \end{cases} \tag{3}$$

where (a,b) are pairs of constraint atoms, which were determined in previous work by calculating all interatomic distances in a number of structures and selecting all pairs that were consistent. The quasi-continuous part of the potential, V_{qc}, uses a sum over harmonic terms around the internal minimum distances in the line segments between atoms in neighbouring interpolation images. While the form of the potential is simple, the location of these local minima is more involved, and a detailed discussion is provided in the original publication.[16]

After the potential is defined, a band of images is constructed, adding one atom at a time to each image, and adjusting the images by minimising V_{aux} using an L-BFGS[57,58] algorithm. If two images are too close they are merged; if they are too far apart another image is added in between. This methodology should ensure a balanced description along the entire path. The interpolation potential is constructed such that its derivatives are smooth and continuous. In contrast to DNEB interpolation, no additional springs were used between the images in the original applications,[16] and no gradient projections were used. The convergence condition is defined using the RMS force on the images, and a failure condition is set if this threshold is not reached within a maximum number of minimisation steps. This framework was shown to be applicable to both biomolecules and atomic clusters bound by isotropic pair potentials.[16]

2.3 Improvements to the QCI algorithm

For the more complex examples considered in the present work, the original QCI formulation did not reliably produce interpolations free of chain crossings. However, with a variety of modifications it has proved possible to produce more robust interpolations involving only physically correct configurations. These interpolations also provide much lower energy initial pathways, and enable converged kinetic transition networks to be characterised in a fraction of the time previously required.

The first modification uses knowledge of the bonded interactions available directly from the AMBER potential setup. Instead of constructing the interpolation potential from constraints determined from a database of local minima, the known bonding topology is exploited. This choice will change the selection of pairs (a, b) in equation (2).

For these more complex tests harmonic springs between the images were found to be beneficial; 'nudging'[11,13] was not required. Additional constraints have also been admitted in a general user-defined array, and this feature was used to add cis-trans restraints via six pairwise distances. A similar approach for chiral centres could be used, and will be tested in future work.

To improve efficiency in obtaining an initial connected path, we allowed interpolations based on either QCI or DNEB, using QCI to bridge larger gaps between conformations, and DNEB for smaller separations. The distances between end points are all minimised with respect to overall translation, rotation, and permutation of identical atoms. [16] These distances are used to define a metric to select pairs of minima for subsequent connection attempts in the OPTIM[59] and PATHSAMPLE[60] programs using a missing connection technique[61,62] based on Dijkstra's shortest path algorithm. [63]

We have also considered optimisation of the QCI parameters in the present work. After an initial screening, where we only changed one parameter at a time, six key parameters were identified: the image recombination threshold, $d_{\mathrm{combine}}^{\mathrm{Images}}$, the maximum number of images allowed, $N_{\mathrm{max}}^{\mathrm{Images}}$, the threshold distance for the activation of the harmonic constraints, r_{attr}, the gradient convergence required before adding another atom, Δ, the cutoff distance for repulsion, r_{rep}, and the spring constant between images, $k_{\mathrm{spr}}^{\mathrm{Images}}$. The second round of testing considered all possible combinations for these six parameters, as shown in Table 1.

Several results were considered to evaluate the efficiency. To judge the quality of the transition states obtained after interpolation we compared the number of candidates, the number of actual transition states found after accurate refinement, the number of rejected transition states based on cis-trans isomerism and chirality inversions, and the number of accepted transition states. Furthermore, we considered the length of the interpolation in Euclidean space as a sum of distances between images, the highest energy image in the band, and the number of images. To test the efficiency the overall time for all calculations for one interpolation cycle, as well as the time to construct the initial set of images were considered. An illustration of the interplay between all these quantities is shown in Fig. 2.

Combinations of these six parameters lead to a set of 4500 possibilities. Additionally, cis-trans constraints were added in a second set of runs with the same 4500 settings. To constrain the peptide bonds six pairwise spring potentials were applied between the two α-carbon atoms on either side of the bond, the carboxyl-O atom, and the H atom on the

Table 1: Summary of the parameter space explored for six key QCI parameters. The remarks supply additional information. The energy units are arbitrary (arbs).

Parameter	Values tested	Remark
$d_{\text{combine}}^{\text{Images}}$ / Å	0.5, 1.0, 1.5, 2.0, 3.0	image recombination threshold
$N_{\text{max}}^{\text{Images}}$	50, 75, 100, 150, 200	the initial number of images is not important as long as the maximum number of images is large enough
r_{attr} / Å	0.05, 0.075, 0.1, 0.125, 0.15	the first round of tests reveales a performance reduction above r_{attr} of 0.2 Å
Δ / arbs Å^{-1}	0.05, 0.1, 0.2, 0.4	RMS gradient convergence required before adding another atom
r_{rep} / Å	2.0, 4.0, 6.0	key parameter: the quasi-continuous penalty terms and the non-bonded interactions both depend on this parameter
$k_{\text{spr}}^{\text{Images}}$ / arbs Å^{-2}	0.1, 0.25, 0.5	non-zero spring constants provided improved performance for more complex interpolations

peptide bond nitrogen.

3 Results

3.1 Optimisation of the parameter space

For most parameters there is a trade off between an increased number of low energy transition states and the run time. For larger systems the interpolation time may become a significant component of the computational cost, and should therefore be as low as possible. Three of the parameters directly influence the timings: the image recombination threshold, the maximum number of images, and the gradient convergence required. Longer, but lower energy, pathways are found for small recombination thresholds and tighter convergence.

An increased maximum number of interpolation images reduces the number of unphysical geometries encountered, which lowers the number of rejected transition states. However, the overall run time increases, as well as the interpolation time. A larger repulsion cutoff increases the required time for both short and long distances. This effect may arise because for short cutoffs there are not enough repulsive interactions, while more numerous interactions increase the time required it evaluate the auxiliary potential. Two parameters, the threshold for activation of the harmonic bond constraints, and the spring constants between images, have little effect on run time. A summary is given in Table 2.

Table 2: Summary of the effects of changing six key parameters on the run time t, and the quality of transition states for an initial connection attempt. In each case the results were averaged over all the other parameters to obtain trends for single parameters.

Parameter	Run Time	Path Properties
$d_{\text{combine}}^{\text{Images}}$	increased t for small values	smaller values lead to lower maximum image energies and more converged transition states
$N_{\text{max}}^{\text{Images}}$	increased t for more more images	no change in the number of transition states found, but for a small number of images more rejections occur
r_{attr}	no significant t dependence	small values (tighter constraints) result in more transition states with lower maximum energies
Δ	increased t required for tighter convergence	decrease in maximum energy for tighter convergence in the candidate structures
r_{rep}	longer t required for small and large values	
$k_{\text{spr}}^{\text{Images}}$	no significant t dependence	for small values more transition states are located

Optimal parameter sets improved the performance significantly, reducing the initial energy of the transition state candidates by up to two orders of magnitude. The converged

transition states generally exhibit energies between 30 to 70 kcal mol^{-1} higher than the endpoints, which gives an initial starting guess in much better agreement with experimentally observed barriers than for standard DNEB interpolations. Furthermore, more of the candidate structures can be converged to genuine transition states. An example can be found in Fig. 3 for amyloid-β_{17-42}.

3.2 The effect of cis-trans constraints

Applying the cis-trans constraints reduces the number of rejections by around 50%. Importantly, the overall time required for the QCI run does not increase, despite the additional constraints considered in the auxiliary potential. The increase in run time for the potential is offset by the reduced time for of transition state refinement. Most of the remaining rejected structures arise from chirality inversions. Additional constraints for chirality would likely reduce this number further, and will be considered in future work. Considering all these trends together we can identify a best parameter set for Aβ_{17-42}, and apply these values for multiple QCI cycles until a complete connected discrete path is located between the two endpoints. In Fig. 4 the best initial path found for the Aβ_{17-42} monomers is shown. We emphasize once more that this initial path is not expected to be the most kinetically relevant for a complex rearrangement. Rather, it provides the starting point for refinement of a kinetic transition network.

3.3 Histone tail example

Using the best parameters identified in the tests for Aβ_{17-42}, and applying them in multiple cycles to the histone tail endpoints, successfully produced a complete connected path. The QCI interpolation was used for all minima separated by more than 10 Å, otherwise an initial linear interpolation was employed with DNEB to locate transition state candidates. The resulting disconnectivity graph[64,65] including all the local minima and transition states is illustrated in Fig. 5.

The initial pathway exhibits no unphysical transition state energies, indicating the absence of any unphysical transitions, as shown in Fig. 6, which is a great improvement over previous unsuccessful DNEB attempts. We then considered additional pairs of local minima for the endpoints without any changes in the simulation parameters. Again, complete initial paths were located without any unphysically high energy barriers. We were not able to find acceptable initial pathways using only DNEB interpolations.

3.4 EGFR kinase

Using QCI for the egfr kinase increased the number of transition states located after the initial interpolation significantly, compared to DNEB interpolations. Additionally, the energies observed for QCI are much lower than for DNEB, in some cases by orders of magnitude. The transition states found using QCI were typically within $70\,\mathrm{kcal\,mol^{-1}}$ of the lower energy endpoint (for comparison, the other endpoint is $20\,\mathrm{kcal\,mol^{-1}}$ higher in energy). Nevertheless, a few problems are encountered for QCI. First, a number of the transition states involve chirality inversion. This observation implies that for larger systems the application of chirality constraints may be necessary. Additionally, the time required for the interpolation becomes quite long, which may be unavoidable in reducing the overall cost of a successful interpolation. At present we have not obtained a full discrete path between the two endpoints, but we will provide a detailed report on this system in future work.

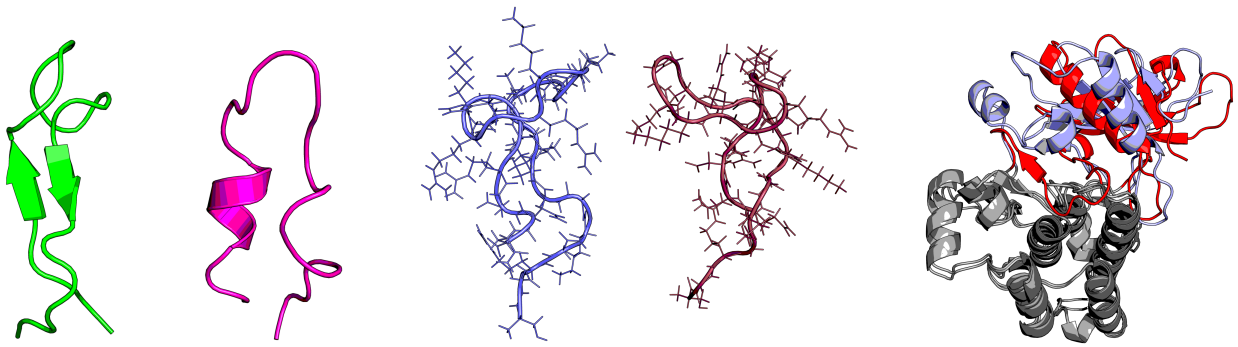
4 Conclusions

We have implemented an improved version of the original quasi-continuous interpolation scheme, incorporating more information known a-priori, and optimising the adjustable parameters. The new version reliably avoids unphysical structures and produces initial paths that are more physically relevant than for DNEB. This approach greatly reduces the computational effort required to obtain converged rates from the resulting kinetic transition

network in the case of Aβ_{17-42} and the histone tail. For the largest test case a similar improvement is observed, with the location of much lower energy transition states. In general, the interpolation images, as well as the converged transition states, are much lower in energy than for standard DNEB runs. The initial paths located generally involve low-energy structures and thermally accessible intermediate states at relevant temperatures. Unphysical transformations, such as chain crossings, as well as unphysical structures, such as inverted cis-trans bonds, are explicitly avoided. It is possible to extend the interpolation potential to prevent chirality inversions or include other constraints between pairs of atoms if necessary. Using QCI for larger gaps, and DNEB for distances below 10 Å, produced significant improvements in efficiency. Hence the new QCI formulation has the potential to characterise complete pathways that have previously been inaccessible. While we have focussed here on the application of QCI within discrete path sampling,[7,8] the initial interpolations could be combined with other two-ended approaches, for example to initialise string methods[26–28] or as non-trivial guesses for interface or milestone locations in transition interface sampling[66] and milestoning.[67]

Acknowledgement

The authors thank Dr Rosana Collepardo-Guevara for providing the histone tail structures, and Dr Joanne Carr for helpful discussions. We gratefully acknowledge the Engineering and Physical Sciences Research Council for financial support.



(a) (b) (c)

Figure 1: (a) α-helical and β-sheet conformations for the Aβ_{17-42} monomer, (b) two of the histone tail configurations, (c) superposition of open and closed egfr kinase conformations with the secondary structure differences highlighted in red and blue. The figures were generated using PyMOL.[68]

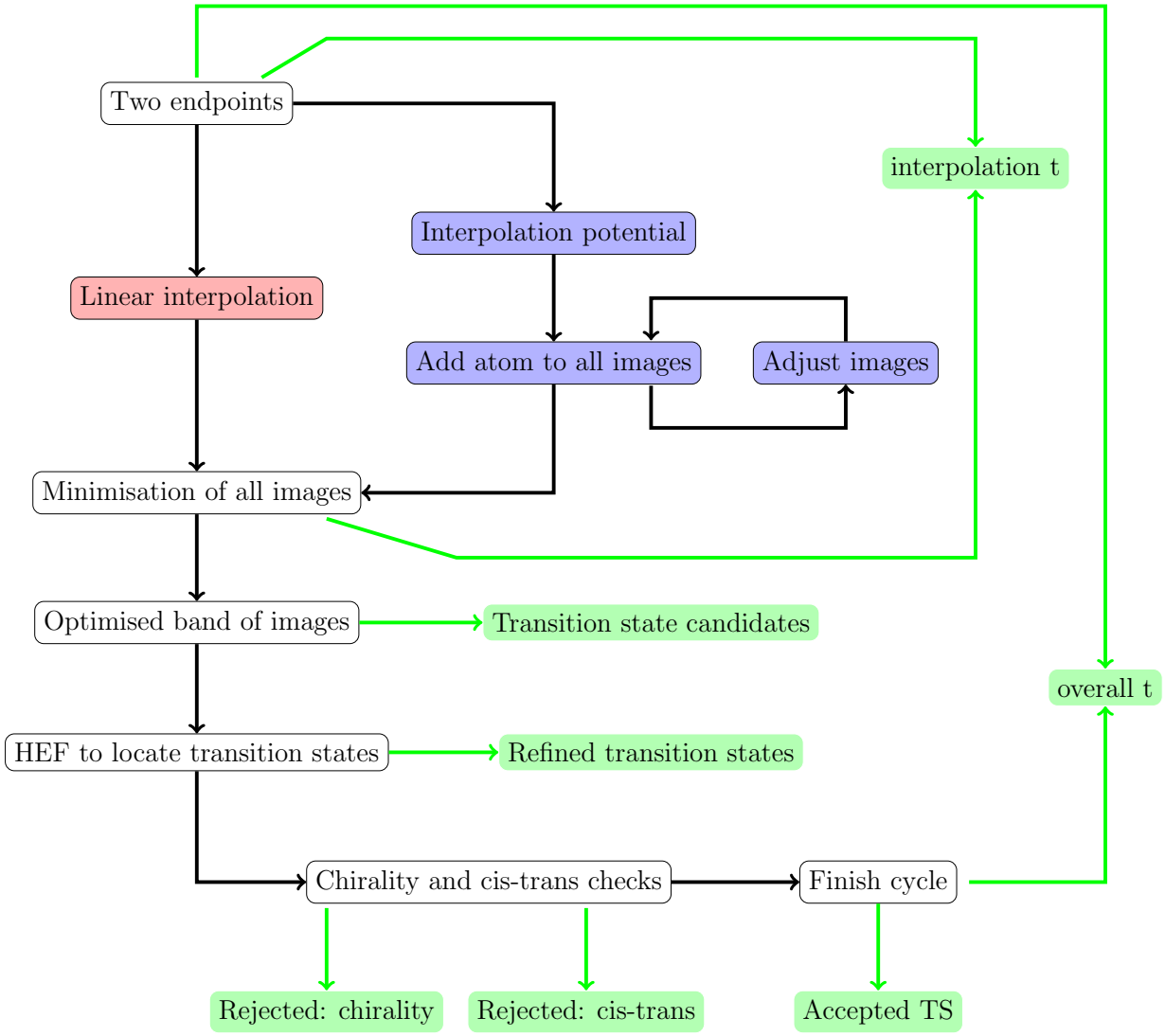


Figure 2: Comparison of the QCI and DNEB interpolation schemes starting from two endpoints. The QCI route is shown in blue and the DNEB equivalent in red. Properties used to test the algorithm are shown in green, and the point where they occur in the algorithm is indicated by arrows. HEF is hybrid eigenvector-following and t is computational time.

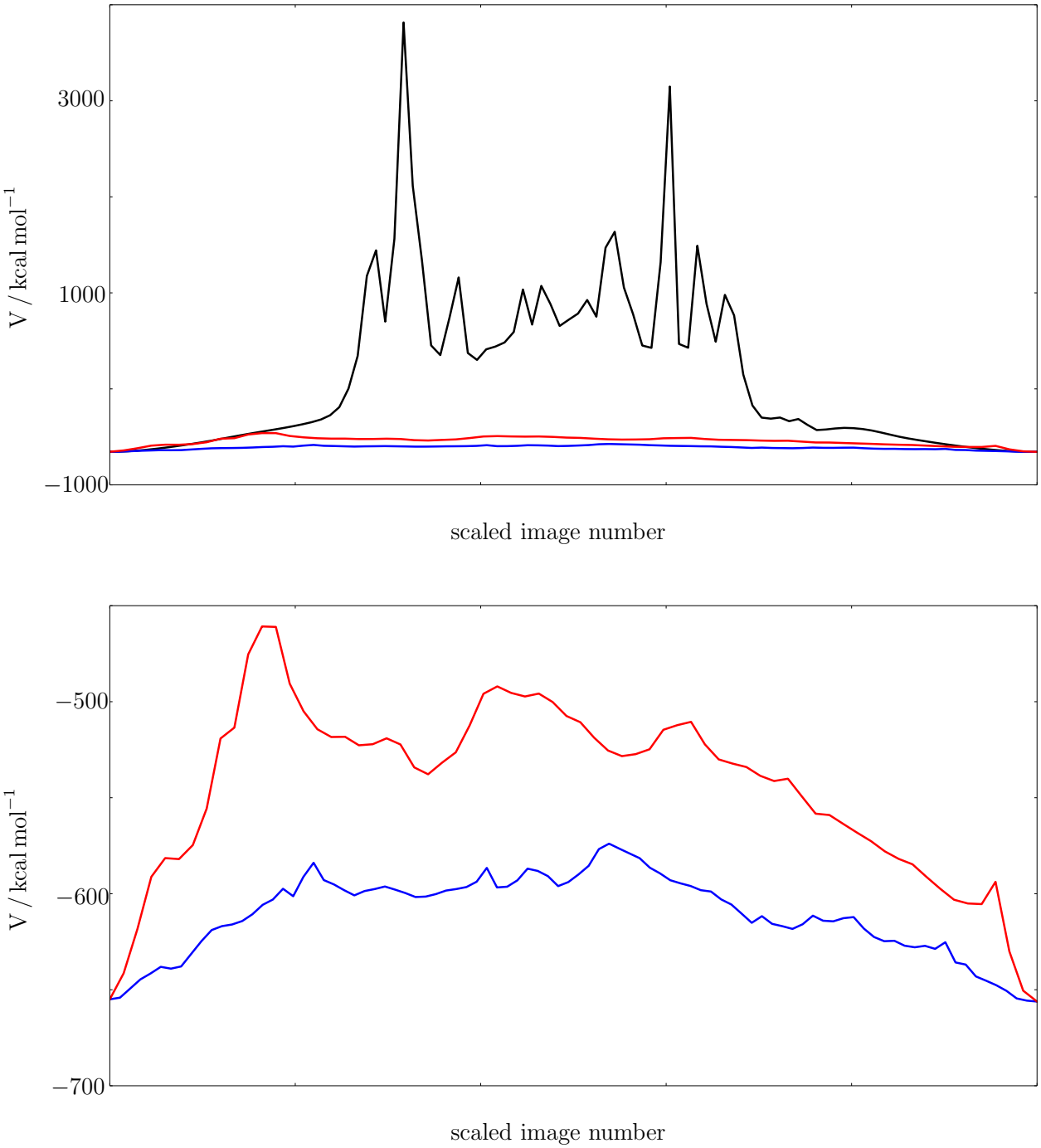


Figure 3: Top: The potential energies of all images for a doubly-nudged elastic band (black) compared to two different QCI runs after 500 iterations of geometry optimisation for each chain of images. Bottom: A closer look at the profile for the two QCI interpolations reveals many candidates for transition states (maxima in the profile) at much lower potential energies than for DNEB with initial linear interpolation (60 kcal mol^{-1} higher than the endpoints for QCI, compared to up to 4500 kcal mol^{-1} for DNEB).

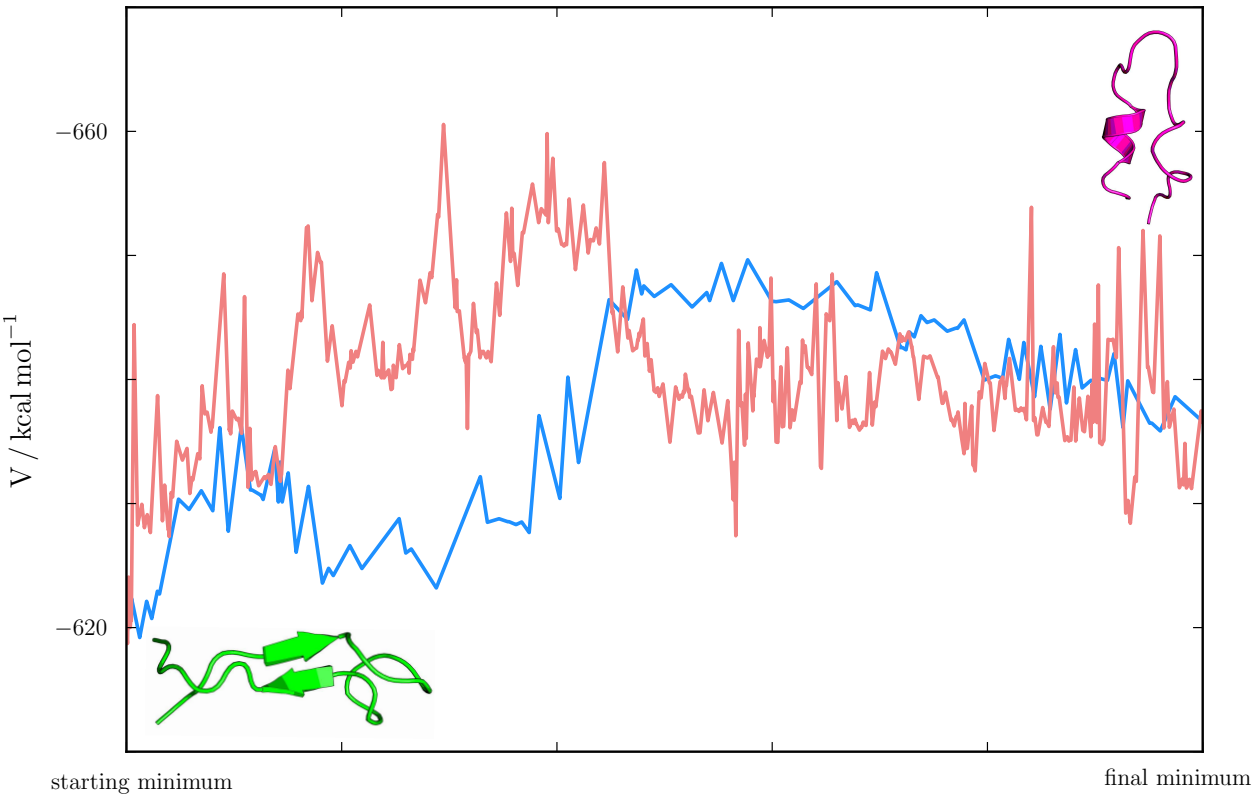


Figure 4: An initial complete discrete path (red) between two endpoints for Aβ_{17-42} is compared to the lowest energy path (blue) found after further exploration of the energy landscape to converge the rates in the resulting kinetic transition network. The initial path is close to the final converged path obtained, and does not exhibit any unphysical energy barriers. The horizontal axis corresponds to the path length of the interpolation, which differs from interpolation to interpolation, and is therefore scaled onto a common axis to allow a more direct comparison.

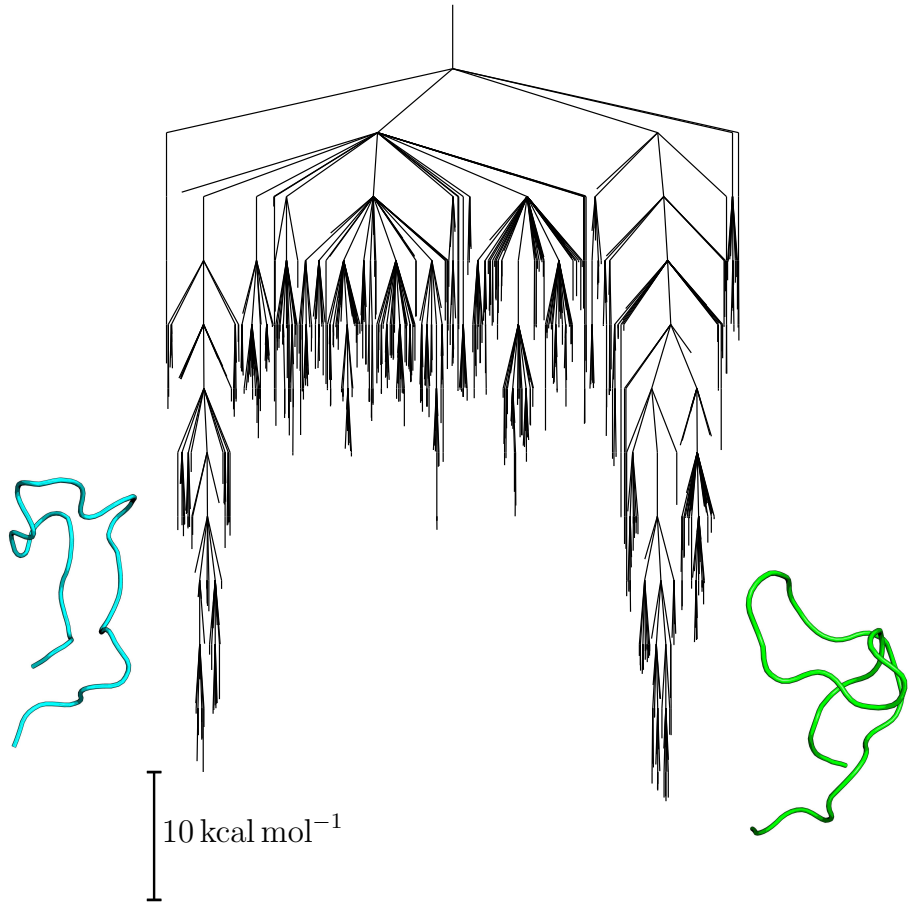


Figure 5: Disconnectivity graph obtained for the transition between two histone tail configurations. The two endpoints are located in the two distinct low energy funnels. The absence of artificially high energy barriers indicates that unphysical geometries have been successfully avoided.

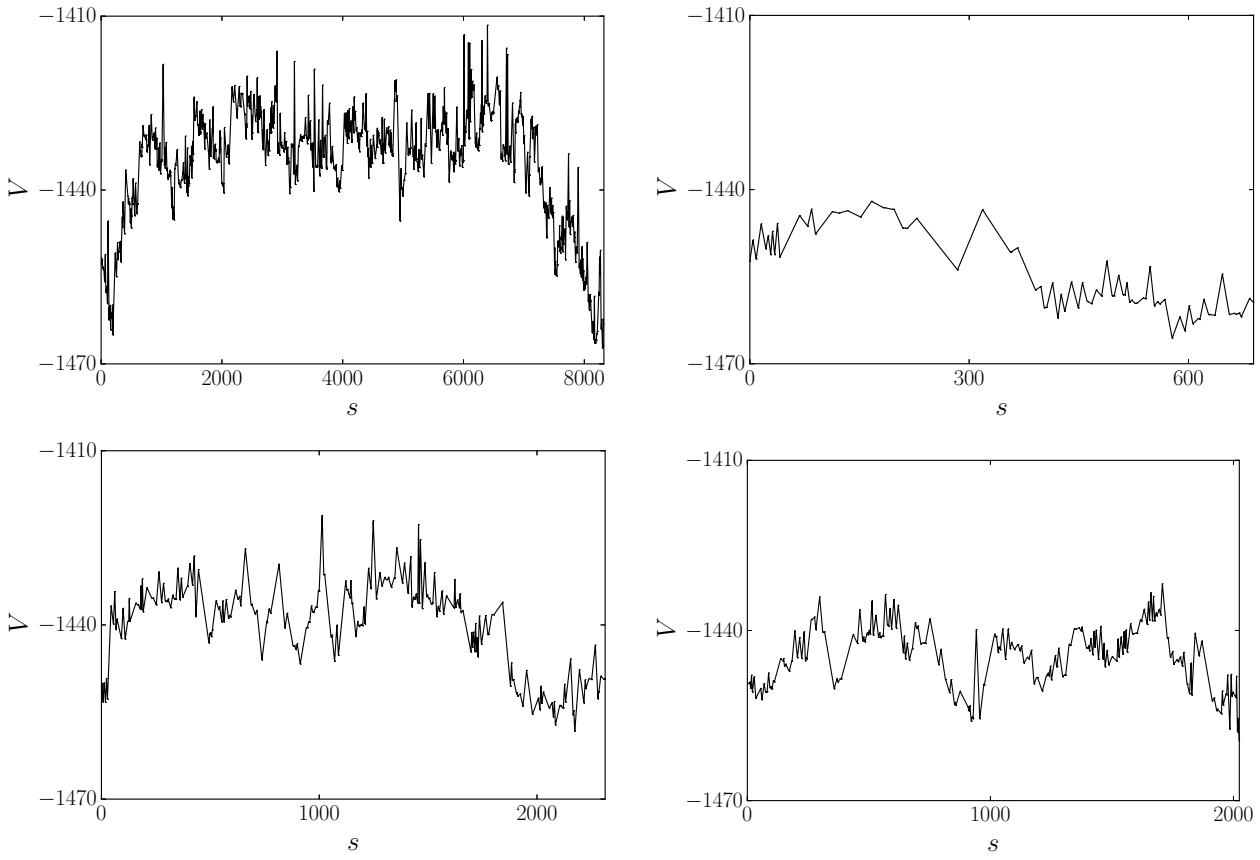


Figure 6: The initial pathways for four different pairs of endpoints for the histone tails exhibit no abnormally high energy barriers, demonstrating that chain crossing, cis-trans isomerism, and chirality inversion have been successfully avoided. V is the potential energy in $\mathrm{kcal\,mol^{-1}}$ and s is the integrated path length in Å. The disconnectivity graph shown in Fig. 5 corresponds to the top right path.

References

(1) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the folding routes. *Science* **1995**, *267*, 1619–1620.

(2) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. Protein folding funnels: the nature of the transition state ensemble. *Fold. Des.* **1996**, *1*, 441–450.

(3) Wales, D. J. *Energy Landscapes*; Cambridge University Press, 2003.

(4) Wolynes, P. G. Energy landscapes and solved protein-folding problems. *Philos. T. Roy. Soc. A* **2005**, *363*, 453–467.

(5) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.* **2008**, *37*, 289–316.

(6) Joseph, J. A.; Röder, K.; Chakraborty, D.; Mantell, R. G.; Wales, D. J. Exploring biomolecular energy landscapes. *Chem. Commun.* **2017**, *53*, 6974–6988.

(7) Wales, D. J. Discrete path sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.

(8) Wales, D. J. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* **2004**, *102*, 891–908.

(9) Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struc. Biol.* **2008**, *18*, 154–162.

(10) Wales, D. J. Energy landscapes: some new horizons. *Curr. Opin. Struc. Biol.* **2010**, *20*, 3–10.

(11) Trygubenko, S. A.; Wales, D. J. A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.* **2004**, *120*, 2082–2094.

(12) Henkelman, G.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(13) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(14) Munro, L. J.; Wales, D. J. Defect migration in crystalline silicon. *Phys. Rev. B* **1999**, *59*, 3969–3980.

(15) Zeng, Y.; Xiao, P.; Henkelman, G. Unification of algorithms for minimum mode optimization. *J. Chem. Phys.* **2014**, *140*, 044115.

(16) Wales, D. J.; Carr, J. M. Quasi-continuous interpolation scheme for pathways between distant configurations. *J. Chem. Theor. Comput.* **2012**, *8*, 5020–5034.

(17) Farrell, D. W.; Speranskiy, K.; Thorpe, M. F. Generating stereochemically acceptable protein pathways. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2908–2921.

(18) Farrell, D. W.; Lei, M.; Thorpe, M. F. Comparison of pathways from the geometric targeting method and targeted molecular dynamics in nitrogen regulatory protein C. *Phys. Biol.* **2011**, *8*, 026017.

(19) Elber, R.; Shalloway, D. Temperature dependent reaction coordinates. *J. Chem. Phys.* **2000**, *112*, 5539–5545.

(20) Faccioli, P. Characterization of protein folding by dominant reaction pathways. *J Phys Chem B* **2008**, *112*, 13756–13764.

(21) Faccioli, P.; Lonardi, A.; Orland, H. Dominant reaction pathways in protein folding: A direct validation against molecular dynamics simulations. *J. Chem. Phys.* **2010**, *133*, 045104.

(22) Faccioli, P.; Beccara, S. Computing reaction pathways of rare biomolecular transitions using atomistic force-fields. *Biophys. Chem.* **2016**, *208*, 62–67.

(23) Wang, F.; Cazzolli, G.; Wintrode, P.; Faccioli, P. Folding mechanism of proteins Im7 and Im9: Insight from all-atom simulations in implicit and explicit solvent. *J. Phys. Chem. B* **2016**, *120*, 9297–9307.

(24) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. Targeted molecular dynamics simulation of conformational change-application to the T R transition in insulin. *Mol. Simul.* **1993**, *10*, 291–308.

(25) Ma, J.; Karplus, M. Molecular switch in signal transduction: Reaction paths of the conformational changes in ras p21. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11905–11910.

(26) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, 052301.

(27) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.* **2004**, *120*, 7877–7886.

(28) Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI. *J. Chem. Phys.* **2011**, *134*, 085103.

(29) Wang, L.-P.; McGibbon, R. T.; Pande, V. S.; Martinez, T. J. Automated discovery and refinement of reactive molecular dynamics pathways. *J. Chem. Theory Comput.* **2016**, *12*, 638–649.

(30) Zhang, S.; Iwata, K.; Lachenmann, M.; Peng, J.; Li, S.; Stimson, E.; Lu, Y.-a.; Felix, A.; Maggio, J.; Lee, J. The Alzheimer's peptide Aβ adopts a collapsed coil structure in water. *J. Struct. Biol.* **2000**, *130*, 130–141.

(31) Glenner, G. G.; Wong, C. W. Alzheimer's disease and Down's syndrome: Sharing of a unique cerebrovascular amyloid fibril protein. *Biochem. Biophys. Res. Commun.* **1984**, *122*, 1131–1135.

(32) Masters, C. L.; Simms, G.; Weinman, N. A.; Multhaup, G.; McDonald, B. L.; Beyreuther, K. Amyloid plaque core protein in Alzheimer disease and Down's syndrome. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 4245–4249.

(33) Selkoe, D. J. Alzheimer's disease is a synaptic failure. *Science* **2002**, *298*, 789–791.

(34) Chiti, F.; Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.

(35) Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N.-V.; Coté, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; Laio, A.; Li, M. S.; Melchionna, S.; Mousseau, N.; Mu, Y.; Paravastu, A.; Pasquali, S.; Rosenman, D. J.; Strodel, B.; Tarus, B.; Viles, J. H.; Zhang, T.; Wang, C.; Derreumaux, P. Amyloid β protein and Alzheimer's disease: When computer simulations complement experimental studies. *Chem. Rev.* **2015**, *115*, 3518–3563.

(36) Nagel-Steger, L.; Owen, M. C.; Strodel, B. An account of amyloid oligomers: Facts and figures obtained from experiments and simulations. *ChemBioChem* **2016**, *17*, 657–676.

(37) Garcia-Ramirez, M.; Dong, F.; Ausio, J. Role of the histone "tails" in the folding of oligonucleosomes depleted of histone H1. *J. Biol. Chem.* **1992**, *267*, 19587–19595.

(38) Shogren-Knaak, M.; Ishii, H.; Sun, J. M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **2006**, *311*, 844–847.

(39) Robinson, P. J.; An, W.; Routh, A.; Martino, F.; Chapman, L.; Roeder, R. G.;

Rhodes, D. 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J. Mol. Biol.* **2008**, *381*, 816–825.

(40) Bannister, A. J.; Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **2011**, *21*, 381–395.

(41) Hynes, N. E.; MacDonald, G. ErbB receptors and signaling pathways in cancer. *Curr. Opin. Cell Biol.* **2009**, *21*, 177–184.

(42) Li, Y.; Li, X.; Ma, W.; Dong, Z. Conformational transition pathways of epidermal growth factor receptor kinase domain from multiple molecular dynamics simulations and Bayesian clustering. *J. Chem. Theory Comput.* **2014**, *10*, 3503–3511.

(43) Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wolf, R.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.; Mathews, D.; Seetin, M.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. AMBER12. University of California, San Francisco, 2012.

(44) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(45) Małolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Symmetrization of the AMBER and CHARMM force fields. *J. Comput. Chem.* **2010**, *31*, 1402–1409.

(46) Małolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Erratum to: Symmetrization of the AMBER and CHARMM force fields. *J. Comput. Chem.* **2012**, *33*, 2209.

(47) Hornak, V.; Abel, R.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.

(48) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(49) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394.

(50) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.

(51) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(52) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(53) Mantell, R. G.; Pitt, C. E.; Wales, D. J. GPU-Accelerated Exploration of Biomolecular Energy Landscapes. *J. Chem. Theory Comput.* **2016**, *12*, 6182–6191.

(54) Collepardo-Guevara, R.; Portella, G.; Vendruscolo, M.; Frenkel, D.; Schlick, T.; Orozco, M. Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study. *J. Am. Chem. Soc.* **2015**, *137*, 10205–10215.

(55) Yun, C.-H.; Boggon, T. J.; Li, Y.; Woo, M. S.; Greulich, H.; Meyerson, M.; Eck, M. J. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **2007**, *11*, 217–227.

(56) Kawakita, Y.; Seto, M.; Ohashi, T.; Tamura, T.; Yusa, T.; Miki, H.; Iwata, H.; Kamiguchi, H.; Tanaka, T.; Sogabe, S.; Ohta, Y.; Ishikawa, T. Design and synthesis of novel pyrimido[4,5-b]azepine derivatives as HER2/EGFR dual inhibitors. *Bioorgan. Med. Chem.* **2013**, *21*, 2250–2261.

(57) Nocedal, J. Updating quasi-newton matrices with limited storage. *Math. Comput.* **1980**, *35*, 773.

(58) Liu, D.; Nocedal, J. On the limited memory bfgs method for large scale optimization. *Math. Prog.* **1989**, *45*, 503.

(59) OPTIM - a program for optimising geometries and calculating reaction pathways, version 33537. Wales, David J.: Cambridge, 2018.

(60) PATHSAMPLE: A driver for OPTIM to create stationary point databases using discrete path sampling and perform kinetic analysis, r33309. Wales, David J.: Cambridge, 2018.

(61) Evans, D. A.; Wales, D. J. Folding of the GB1 hairpin peptide from discrete path sampling. *J. Chem. Phys.* **2004**, *121*, 1080–1090.

(62) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. Finding pathways between distant local minima. *J. Chem. Phys.* **2005**, *122*, 234903.

(63) Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Mat.* **1959**, *1*, 269–271.

(64) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces:

Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **1997**, *106*, 1495–1517.

(65) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal energy landscapes. *Nature* **1998**, *394*, 758–760.

(66) van Erp, T. S.; Moroni, D.; Bolhuis, P. G. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* **2003**, *118*, 7762–7774.

(67) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **2004**, *120*, 10880–10889.

(68) The PyMOL Molecular Graphics System, Version 1.6.x. Schrödinger, LLC: New York, 2013.

Graphical TOC Entry